Dados de Proveniência como Cidadãos de Primeira Classe para Análises de *Workflows* em Aprendizado Profundo[1]

Débora Pina[1], Daniel de Oliveira[2], Marta Mattoso[1]

[1]PESC/COPPE, Universidade Federal do Rio de Janeiro (UFRJ)
[2]Instituto de Computação, Universidade Federal Fluminense (UFF)

Resumo. A adoção de modelos de Aprendizado Profundo (AP), na tomada de decisões requer confiança e interpretação por parte de usuários do modelo. A proveniência surge como uma solução natural para promover rastros do *workflow* de AP, passíveis de análise, englobando desde a preparação de dados ao modelo de AP. Apesar de diversas abordagens alegarem prover proveniência, a principal limitação está na ausência de mecanismos que possibilitem a análise do caminho de derivação dos dados após a geração do resultado do *workflow* de AP. As soluções existentes não oferecem a rastreabilidade do *workflow*, adotam formatos proprietários para a representação dos metadados e não geram documentos de proveniência que acompanhem os modelos no ambiente de produção. A DLProv é um conjunto de serviços de proveniência que resolve questões de captura e interoperabilidade de rastros de proveniência de diferentes etapas do *workflow*, independente do ambiente de AP. Os serviços geram grafos de proveniência aderentes ao W3C PROV, que contempla as etapas executadas no *workflow* de AP. A DLProv foi avaliada em ambientes de execução de alto desempenho, explorando casos de uso heterogêneos. Os grafos podem ser consultados pelos usuários do modelo em produção independente do acesso aos ambientes de geração do modelo.

Abstract. The adoption of Deep Learning (DL) models in decision-making requires trust and interpretation for the users of the model. Provenance emerges as a natural solution for generating traces of the DL workflow, which can be analyzed and span from data preparation to the DL model. Although several approaches claim to provide provenance, the main limitation is the lack of mechanisms that enable analysis of the data derivation path after the DL workflow results are generated. Existing solutions lack DL workflow traceability, adopt proprietary formats for representing metadata, and do not generate provenance documents that accompany the models in the production environment. DLProv is a suite of provenance services that supports the capture and interoperability of provenance traces across different workflow stages, independent of the DL environment. The services generate provenance graphs compliant with W3C PROV, which include the steps executed in the DL workflow. DLProv has been evaluated in high-performance computing environments, using heterogeneous use cases. The graphs can be queried by users of the production model regardless of access to the DL model generation environments.

1. Dados da Defesa da Tese e Pontos de Destaque

Data de Defesa: 24/03/2025
Programa de Pós-Graduação: Engenharia de Sistemas e Computação, COPPE
Universidade: Universidade Federal do Rio de Janeiro (UFRJ)
Categoria: Doutorado
Orientadora: Marta Lima de Queirós Mattoso (PESC/COPPE/UFRJ)
Coorientador: Daniel Cardoso Moraes de Oliveira (IC/UFF)
Doutorado Sanduíche: University of Southampton, sob supervisão de Adriane Chapman
Membros da Banca: Adriane Chapman – University of Southampton
Aline Marins Paes Carvalho – Universidade Federal Fluminense (UFF)
Altigran Soares da Silva – Universidade Federal do Amazonas (UFAM)
Claudio Miceli de Farias – Universidade Federal do Rio de Janeiro (COPPE/UFRJ)
Fábio André Machado Porto – Laboratório Nacional de Computação Científica (LNCC)
Nota e Menção Honrosa: Não se aplica
Highlights:

(1) Bases de proveniência para análise de *workflows* de Aprendizado Profundo (AP);
(2) Coleta de proveniência de *workflows* de AP com baixa sobrecarga computacional;
(3) Serviços de proveniência invocados por *scripts*, bibliotecas de AP ou embarcados;
(4) Grafos de proveniência validados em conformidade com W3C PROV;
(5) Exporta o grafo de proveniência de um modelo de AP em múltiplos formatos.

[1]*Provenance Data as a First-Class Citizen for Deep Learning Workflow Analyses*

2. Contexto e Problema

Um modelo de aprendizado profundo (AP) é projetado para analisar dados e gerar resultados preditivos. Sua construção é um processo iterativo que envolve avaliação de métricas e ajustes sucessivos [9], e segue um *workflow* com etapas como carregamento, preparação (incluindo divisão) dos dados, treinamento e avaliação do modelo [10]. Após gerar modelos candidatos, seleciona-se o melhor com base em critérios relevantes ao usuário, e este é implantado em produção [19]. A confiabilidade do modelo, necessária para promover uma tomada de decisão de qualidade [43], depende da rastreabilidade das etapas envolvidas em sua geração, sendo importante manter registros dos dados, transformações aplicadas ao longo do *workflow* e agentes envolvidos [36, 19]. No entanto, a análise da literatura mostra que esse registro é incipiente, levando a análises ambíguas e pouco confiáveis, pois dependem de integrações manuais de dados e do desenvolvimento de programas para correlacionar etapas, algo muitas vezes inviável, especialmente em ambientes de produção. Dessa forma, a análise integrada torna-se desafiadora quando os dados estão registrados de forma desconectada ou sem representar relacionamentos passíveis de consultas [15, 21]. O principal problema das soluções existentes é que elas mesclam os dados usados para apoiar o cientista de dados com aqueles que permitiriam a geração dos grafos que promovem a rastreabilidade, o que compromete a representação dos relacionamentos entre etapas, introduz ruído com dados irrelevantes à rastreabilidade e gera dependência da ferramenta de modelagem, muitas vezes inacessível após a implantação. Até onde sabemos, não há soluções de rastreabilidade que abranjam todo o ciclo de vida do modelo, o que compromete a análise integrada das etapas que levaram ao modelo escolhido [24, 29], e, por consequência, afeta a transparência, auditabilidade e confiabilidade dos sistemas de AP. Por exemplo, consultas como *"Quais filtros foram aplicados na preparação dos dados de um modelo que atingiu acurácia superior a x?"* não são possíveis ou são sujeitas a interpretações subjetivas.

3. Objetivo

A proveniência é uma solução natural para oferecer rastreabilidade [24], pois, por meio da captura das proveniências prospectiva e retrospectiva [16, 18], permite registrar tanto a estrutura das etapas necessárias e a abstração do fluxo de dados, representando a sequência de tarefas encadeadas por transformações, conjuntos de dados e suas dependências, quanto os dados sobre a execução do *workflow*. O objetivo desta tese é oferecer a proveniência como cidadão de primeira classe no contexto de AP, promovendo uma abordagem modular, estruturada e independente de *frameworks* de AP para capturar, gerenciar e analisar dados de proveniência do ciclo de vida dos modelos de AP. Nossa hipótese é que *Proveniência como Serviço* é essencial para fornecer a rastreabilidade necessária em *workflows* de AP. Para alcançar esse objetivo, esta tese apresenta a DLProv.

4. Contribuição

DLProv contribui com uma abordagem centrada em proveniência para prover rastreabilidade em *workflows* de AP dando flexibilidade para o cientista de dados escolher as melhores ferramentas para cada etapa do *workflow* sem prejuízo de obtenção do grafo de proveniência. Do ponto de vista científico, as contribuições incluem *(i)* a proposição e implementação de uma arquitetura baseada no conceito de Proveniência como Serviço, capaz de se integrar com diferentes ferramentas e *workflows*; *(ii)* a definição de um modelo de dados compatível com o W3C PROV, assegurando a representação explícita de relacionamentos típicos de *workflows* de AP; *(iii)* a geração de documentos de proveniência do modelo com o rastreio das etapas de seu *workflow*; e *(iv)* os cenários reais de uso da DLProv, que evidenciaram a importância dos grafos de proveniência ao serem empacotados junto ao modelo para uso de terceiros, promovendo confiança e interpretação em ambientes isolados dos *frameworks* que geraram o modelo. A DLProv é disruptiva ao prover um grafo de proveniência autônomo, validado pelo padrão como um serviço *cidadão de primeira classe* de baixa sobrecarga computacional. DLProv viabiliza a geração desses grafos independentes seja ao ser invocada em *scripts* [28, 29, 40], ou acoplada em bibliotecas [14], ou ainda embarcada em *frameworks* [31, 33]. O código-fonte da DLProv está em: `https://github.com/dbpina/dlprov`.

5. Avanço no Estado da Arte

A revisão da literatura mostra que as soluções existentes para rastreabilidade em AP concentram-se na gestão de metadados em nível de entidade [48, 12, 17, 44, 22, 23, 26, 37], negligenciando a representação explícita dos relacionamentos entre atividades, agentes, e entidades dos caminhos de derivação de dados. Além disso, a ausência de padronização na representação de dados é uma limitação recorrente nas soluções existentes [34, 47, 38, 17, 44, 26, 27], que adotam representação *ad-hoc*. Essa limitação se estende a ferramentas amplamente utilizadas, como Comet [1], MLflow [48, 12, 3], e Weights and Biases [8], que utilizam formatos proprietários, dificultando a interoperabilidade entre diferentes ambientes e plataformas. Algumas dessas ferramentas se restringem a linguagens específicas, como Python [44, 26], ou a *frameworks* de Aprendizado de Máquina (AM) [35, 46, 45]. Mesmo soluções mais flexíveis, como [42, 27], exigem sua incorporação em todas as etapas do

workflow, o que limita a integração com outras soluções de captura de proveniência. As soluções existentes não fornecem um documento de proveniência independente que possa acompanhar o modelo após sua implantação. Uma exceção parcial é o MLflow2PROV [37], que, no entanto, depende do MLflow. Essa limitação evidencia uma lacuna ainda não resolvida na literatura, que gera uma dependência do ambiente original de execução, dificultando análises posteriores, verificação e auditoria de modelos em produção. Esta tese se destaca, até onde sabemos, por ser a única a propor uma abordagem de rastreabilidade que representa, de forma padronizada, tanto os metadados (entidades) quanto os relacionamentos entre as atividades, agentes, e entidades do *workflow* de AP, promovendo interoperabilidade, rastreabilidade e independência do ambiente de execução. Com isso, avança o estado da arte ao viabilizar uma análise que abrange o ciclo de vida, independente da execução integral do *workflow* de AP em um único arcabouço e contribui para a confiabilidade dos modelos implantados, promovendo transparência com rastreabilidade [29, 30].

6. Resumo da Solução

DLProv [28, 29, 32] aborda as lacunas mencionadas, tratando a rastreabilidade como um cidadão de primeira classe. A DLProv baseia-se em trabalhos anteriores, como a DfAnalyzer [41] e oferece serviços de proveniência para capturar e exportar proveniências prospectiva e retrospectiva de forma independente de *frameworks* de AP, através da instrumentação de *scripts*. DLProv adota a recomendação W3C PROV [25] para representar dados de proveniência [2], promovendo a interoperabilidade entre sistemas. Os grafos de proveniência da DLProv incluem nós de atividade, agente e entidade, que representam valores de dados, transformações realizadas durante o *workflow* de AP, os atores responsáveis por essas transformações e as máquinas nas quais foram executadas. A DLProv permite a geração e exportação de documentos de proveniência em formatos como JSON e W3C PROV-N, utilizando ferramentas como ProvPy [5], que facilitam a manipulação e conversão desses documentos. Esses documentos podem ser ingeridos em bancos de dados de grafos por meio de ferramentas como o prov2neo [7] e PROV Database Connector [6]. Por exemplo, a DLProv utiliza esse último para permitir que os usuários salvem documentos PROV no Neo4j [4]. Embora se baseie em conceitos consolidados como o W3C PROV e dialogue com serviços como a DfAnalyzer, esta tese propõe uma arquitetura de geração de grafos de proveniência voltada especificamente para etapas dos *workflows* de AP executadas de modo independente e em ambientes reais, rompendo com o paradigma de proveniência acoplada à plataforma de execução e abrindo caminho para a portabilidade e longevidade da rastreabilidade dos modelos.

7. Avaliação

A DLProv foi avaliada quanto a *(a)* integração com outras soluções de proveniência; *(b)* independência de *framework*; *(c)* independência de ambientes de computação; *(d)* poder de análise; e *(e)* exportação de documentos de proveniência. Para avaliar a arquitetura de serviços da DLProv, realizamos sua invocação explicitamente em *scripts* instrumentados, via *callbacks* e encapsulado no Keras. Essas formas de uso da DLProv ressaltam a adequação e importância de serviços de proveniência como cidadão de primeira classe. Para validar a rastreabilidade, compilamos consultas da literatura, do Data Science Stack Exchange e adaptadas do Provenance Challenges [28, 29], uma vez que atualmente não há *benchmark* para esse fim. Os resultados mostram que a DLProv é capaz de capturar e integrar dados de proveniência ao longo do *workflow*, fornecendo uma visão conectada das etapas, artefatos e atores envolvidos no modelo de AP. Para avaliar *(a)*, conduzimos experimentos que integraram a proveniência do pré-processamento, capturada pela DPDS [11] com a proveniência gerada pela DLProv durante o treinamento de modelos de AP [28]. Utilizamos dados públicos, como Framingham Heart Disease, Adult Census Income e Credit Card Fraud Detection. Para avaliar *(b)* e *(c)*, realizamos experimentos com diferentes *frameworks* de AP, incluindo TensorFlow, Keras, PyTorch e DeepXDE [30, 13, 14, 33, 39, 40, 20, 31]. Esses experimentos mostraram sua flexibilidade sendo executados em ambientes computacionais heterogêneos, como computadores pessoais, máquinas com múltiplas GPUs, e ambientes de alto desempenho, como Lobo Carneiro, Grid5000, e Santos Dumont, além de plataformas como Google Colab. Para comparar com o estado da prática e avaliar *(d)*, comparamos a rastreabilidade da DLProv com os recursos do Weights and Biases, MLflow e MLflow2PROV [29]. Esses experimentos mostraram o poder de análise das consultas com a DLProv e as limitações significativas das demais soluções. Para avaliar *(e)*, realizamos experimentos simulando a implantação um modelo AP e geramos um grafo de proveniência desse modelo, exportando-o em formato PROV-N, ingerido no Neo4j [30]. Mostramos que, mesmo após a implantação, é possível reconstruir e analisar seu histórico com base nesses documentos, o que confere autonomia e auditabilidade ao processo. Além disso, nos experimentos realizados, a sobrecarga introduzida pela captura de proveniência com a DLProv foi de, no máximo, 4%. O impacto prático foi verificado nos diversos estudos de caso, incluindo os *workflows* reais. No caso dos modelos de AP guiados pela Física para imageamento sísmico, foi possível acessar e consultar os documentos de proveniência completos mesmo após perder o acesso à infraestrutura original de execução, permitindo a análise detalhada de quais dados e parâmetros foram utilizados na execução para identificar possíveis divergências em produção.

8. Produção Técnica e Científica

Artigos Diretamente Resultantes da Tese

- **Débora Pina**, Liliane Kunstmann, Adriane Chapman, Daniel de Oliveira, and Marta Mattoso. DLProv: a Suite of Provenance Services for Deep Learning Workflow Analyses. Journal PeerJ Computer Science 11:e2985, 2025. DOI: 10.7717/peerj-cs.2985.
- **Débora Pina**, Adriane Chapman, Liliane Kunstmann, Daniel de Oliveira, and Marta Mattoso. DLProv: A Data-Centric Support for Deep Learning Workflow Analyses. In Proceedings of the Eighth Workshop on Data Management for End-to-End Machine Learning (DEEM '24), colocated with ACM SIGMOD. ACM, pages 77–85, 2024. DOI: 10.1145/3650203.3663337. Best Paper Award.
- **Débora Pina**, Adriane Chapman, Daniel De Oliveira, and Marta Mattoso. Deep learning provenance data integration: A practical approach. In Provenance Week 2023, Companion Proceedings of the ACM Web Conference 2023, WWW'23 Companion, pages 1542–1550, 2023. ACM. DOI: 10.1145/3543873.3587561.
- **Débora Pina**, Liliane Kunstmann, Daniel de Oliveira, and Marta Mattoso. Breadcrumbs for your deep learning model: Following provenance traces with DLProv. Journal Software Impacts, 23:100730, 2025. DOI: 10.1016/j.simpa.2024.100730.
- **Débora Pina**, Liliane Kunstmann, Felipe Bevilaqua, Isabela Siqueira, Alan Lyra, Daniel de Oliveira, and Marta Mattoso. Capturing provenance from deep learning applications using keras-prov and colab: a practical approach. Journal of Information and Data Management, 13(5), 2022. DOI: 10.5753/jidm.2022.2544.
- **Débora Pina**, Liliane Neves, Daniel de Oliveira, and Marta Mattoso. Captura automática de dados de proveniência de experimentos de aprendizado de máquina com keras-prov. In Anais Estendidos do XXXVI Simpósio Brasileiro de Bancos de Dados, pages 69–74, 2021. SBC. DOI: 10.5753/sbbd_estendido.2021.18165.
- Filipe Silva, **Débora Pina**, Liliane Kunstmann, and Marta Mattoso. Painel de proveniência: análise durante o treinamento de redes neurais profundas. In Anais Estendidos do XXXVI Simpósio Brasileiro de Bancos de Dados, pages 22–28, 2021. SBC. DOI: 10.5753/sbbd_estendido.2021.18158.
- Rômulo Silva, **Débora Pina**, Liliane Kunstmann, Daniel de Oliveira, Patrick Valduriez, Alvaro Coutinho, and Marta Mattoso. Capturing provenance to improve the model training of pinns: first hand-on experiences with Grid5000. In CILAMCE-PANACM, pages 1–7, 2021.
- Liliane Kunstmann, **Débora Pina**, Filipe Silva, Aline Paes, Patrick Valduriez, Daniel de Oliveira, and Marta Mattoso. Online deep learning hyperparameter tuning based on provenance analysis. Journal of Information and Data Management, 12(5), 2021. DOI: 10.5753/jidm.2021.1924.

Extensão da DLProv em Parceria com a Dissertação de Lyncoln S. de Oliveira

- Lyncoln S. de Oliveira, Liliane Kunstmann, **Débora Pina**, Daniel de Oliveira, and Marta Mattoso. PINN-Prov: Provenance for physics-informed neural networks. In 2023 International Symposium on Computer Architecture and High Performance Computing Workshops (SBAC-PADW), pages 16–23, 2023. DOI: 10.1109/SBAC-PADW60351.2023.00013.
- Lyncoln de Oliveira, Rômulo Silva, Liliane Kunstmann, **Débora Pina**, Daniel de Oliveira, Alvaro Coutinho, and Marta Mattoso. Dados de proveniência para redes neurais guiadas pela física: o caso da equação eikonal. In Anais do XXXVII Simpósio Brasileiro de Bancos de Dados, pages 373–378, 2022. SBC. DOI: 10.5753/sbbd.2022.225367.

Produção Técnica, Registros e Patentes

- **DLProv**: DOI: 10.5281/zenodo.15272133
  SWHID: swh:1:dir:913abb52492ebeef6a81c63992abf92a7c7f4b1e

9. Prêmios e Financiamentos

- Prêmio de Melhor Artigo Longo recebido no *Workshop on Data Management for End-to-End Machine Learning* (DEEM), realizado em conjunto com o SIGMOD, 2024.
- Prêmio de Melhor Artigo de Demonstração recebido no Simpósio Brasileiro de Banco de Dados (SBBD), 2021.
- Financiamento para a *ACM Europe Summer School* em Computação de Alto Desempenho no Centro de Supercomputação de Barcelona[2], 2024.
- Financiamento para participação no *Workshop on Data Management for End-to-End Machine Learning* (DEEM), 2024.
- Bolsa de Doutorado (GD) CNPq.
- Bolsa de Doutorado Sanduíche pela CAPES-PrInt, na *University of Southampton*, Reino Unido.

[2] https://europe.acm.org/seasonal-schools/hpc/2024

Referências

[1] Comet. https://www.comet.com/. Accessed: 2025-06-03.

[2] Dlprov provenance data model. https://github.com/dbpina/dlprov/blob/main/ProvenanceDataModel. Accessed: 2025-06-03.

[3] Mlflow. https://mlflow.org/. Accessed: 2025-06-03.

[4] Neo4j. https://neo4j.com. Accessed: 2025-06-03.

[5] Prov. https://pypi.org/project/prov/. Accessed: 2025-06-03.

[6] Prov database connector. https://github.com/DLR-SC/prov-db-connector. Accessed: 2025-06-03.

[7] prov2neo. https://github.com/DLR-SC/prov2neo. Accessed: 2025-06-03.

[8] Weights and biases. https://wandb.ai/site/. Accessed: 2025-06-03.

[9] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. Software engineering for machine learning: A case study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, pages 291–300, 2019.

[10] Matthias Boehm, Arun Kumar, and Jun Yang. Data management in machine learning systems. *Synthesis Lectures on Data Management*, 11(1):1–173, 2019.

[11] Adriane Chapman, Paolo Missier, Giulia Simonelli, and Riccardo Torlone. Capturing and querying fine-grained provenance of preprocessing pipelines in data science. *Proceedings of the VLDB Endowment*, 14(4):507–520, 2020.

[12] Andrew Chen, Andy Chow, Aaron Davidson, Arjun DCunha, Ali Ghodsi, Sue Ann Hong, Andy Konwinski, Clemens Mewald, Siddharth Murching, Tomas Nykodym, Paul Ogilvie, Mani Parkhe, Avesh Singh, Fen Xie, Matei Zaharia, Richard Zang, Juntai Zheng, and Corey Zumar. Developments in mlflow: A system to accelerate the machine learning lifecycle. DEEM'20. ACM, 2020.

[13] Lyncoln de Oliveira, Rômulo Silva, Liliane Kunstmann, Débora Pina, Daniel de Oliveira, Alvaro Coutinho, and Marta Mattoso. Dados de proveniência para redes neurais guiadas pela física: o caso da equação eikonal. In *Anais do XXXVII Simpósio Brasileiro de Bancos de Dados*, pages 373–378, Porto Alegre, RS, Brasil, 2022. SBC.

[14] Lyncoln S. de Oliveira, Liliane Kunstmann, Débora Pina, Daniel de Oliveira, and Marta Mattoso. Pinnprov: Provenance for physics-informed neural networks. In *2023 International Symposium on Computer Architecture and High Performance Computing Workshops (SBAC-PADW)*, pages 16–23, 2023.

[15] Rafael Ferreira da Silva, Deborah Bard, Kyle Chard, Shaun DeWitt, Ian T. Foster, Tom Gibbs, Carole Goble, William Godoy, Johan Gustafsson, Utz-Uwe Haus, Stephen Hudson, Shantenu Jha, Laila Los, Drew Paine, Frederic Suter, Logan Ward, Sean Wilkinson, Marcos Amaris, Yadu Babuji, Jonathan Bader, Riccardo Balin, Daniel Balouek, Sarah Beecroft, Khalid Belhajjame, Rajat Bhattarai, Wes Brewer, Paul Brunk, Silvina Caino-Lores, Henri Casanova, Daniela Cassol, Jared Coleman, Taina Coleman, Iacopo Colonnelli, Anderson Andrei Da Silva, Daniel de Oliveira, Pascal Elahi, Nour Elfaramawy, Wael Elwasif, Brian Etz, Thomas Fahringer, Wesley Ferreira, Rosa Filgueira, Jacob Fosso Tande, Luiz Gadelha, Andy Gallo, Daniel Garijo, Yiannis Georgiou, Philipp Gritsch, Patricia Grubel, Amal Gueroudji, Quentin Guilloteau, Carlo Hamalainen, Rolando Hong Enriquez, Lauren Huet, Kevin Hunter Kesling, Paula Iborra, Shiva Jahangiri, Jan Janssen, Joe Jordan, Sehrish Kanwal, Liliane Kunstmann, Fabian Lehmann, Ulf Leser, Chen Li, Peini Liu, Jakob Luettgau, Richard Lupat, Jose M. Fernandez, Ketan Maheshwari, Tanu Malik, Jack Marquez, Motohiko Matsuda, Doriana Medic, Somayeh Mohammadi, Alberto Mulone, John-Luke Navarro, Kin Wai Ng, Klaus Noelp, Bruno P. Kinoshita, Ryan Prout, Michael R. Crusoe, Sashko Ristov, Stefan Robila, Daniel Rosendo, Billy Rowell, Jedrzej Rybicki, Hector Sanchez, Nishant Saurabh, Sumit Kumar Saurav, Tom Scogland, Dinindu Senanayake, Woong Shin, Raul Sirvent, Tyler Skluzacek, Barry Sly-Delgado, Stian Soiland-Reyes, Abel Souza, Renan Souza, Domenico Talia, Nathan Tallent, Lauritz Thamsen, Mikhail Titov, Ben Tovar, Karan Vahi, Eric Vardar-Irrgang, Edite Vartina, Yuandou Wang, Merridee Wouters, Qi Yu, Ziad Al Bkhetan, and Mahnoor Zulfiqar. Workflows community summit 2024: Future trends and challenges in scientific workflows. (ORNL/TM-2024/3573), 2024.

[16] Juliana Freire, David Koop, Emanuele Santos, and Cláudio T. Silva. Provenance for computational tasks: A survey. *Computing in Science & Engineering*, 10(3):11–21, 2008.

[17] Gharib Gharibi, Vijay Walunj, Raju Nekadi, Raj Marri, and Yugyung Lee. Automated end-to-end management of the modeling lifecycle in deep learning. *Empirical Software Engineering*, 26:1–33, 2021.

[18] Melanie Herschel, Ralf Diestelkämper, and Houssem Ben Lahmar. A survey on provenance: What for? what form? what from? *The VLDB Journal*, 26:881–906, 2017.

[19] Samuel Idowu, Osman Osman, Daniel Strüber, and Thorsten Berger. Machine learning experiment management tools: a mixed-methods empirical study. *Empirical Software Engineering*, 29(4):1–35, 2024.

[20] Liliane Kunstmann, Débora Pina, Filipe Silva, Aline Paes, Patrick Valduriez, Daniel de Oliveira, and Marta Mattoso. Online deep learning hyperparameter tuning based on provenance analysis. *Journal of Information and Data Management*, 12(5), Nov. 2021.

[21] Simone Leo, Michael R. Crusoe, Laura Rodríguez-Navas, Raül Sirvent, Alexander Kanitz, Paul De Geest, Rudolf Wittner, Luca Pireddu, Daniel Garijo, José M. Fernández, Iacopo Colonnelli, Matej Gallo, Tazro Ohta, Hirotaka Suetake, Salvador Capella-Gutierrez, Renske de Wit, Bruno P. Kinoshita, and Stian Soiland-Reyes. Recording provenance of workflow runs with ro-crate. *PLOS ONE*, 19(9):1–35, 09 2024.

[22] Hui Miao, Ang Li, Larry S Davis, and Amol Deshpande. Modelhub: Towards unified data and lifecycle management for deep learning. *arXiv preprint arXiv:1611.06224*, 2016.

[23] Hui Miao, Ang Li, Larry S. Davis, and Amol Deshpande. Modelhub: Deep learning lifecycle management. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pages 1393–1394, 2017.

[24] Marçal Mora-Cantallops, Salvador Sánchez-Alonso, Elena García-Barriocanal, and Miguel-Angel Sicilia. Traceability for trustworthy ai: A review of models and tools. *Big Data and Cognitive Computing*, 5(2):20, 2021.

[25] Luc Moreau and Paul Groth. Provenance: an introduction to prov. *Synthesis Lectures on the Semantic Web: Theory and Technology*, 3(4):1–129, 2013.

[26] Mohammad Hossein Namaki, Avrilia Floratou, Fotis Psallidas, Subru Krishnan, Ashvin Agrawal, Yinghui Wu, Yiwen Zhu, and Markus Weimer. Vamsa: Automated provenance tracking in data science scripts. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '20, page 1542–1551, New York, NY, USA, 2020. Association for Computing Machinery.

[27] David Nigenda, Zohar Karnin, Muhammad Bilal Zafar, Raghu Ramesha, Alan Tan, Michele Donini, and Krishnaram Kenthapadi. Amazon sagemaker model monitor: A system for real-time insights into deployed machine learning models. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3671–3681, 2022.

[28] Débora Pina, Adriane Chapman, Daniel De Oliveira, and Marta Mattoso. Deep learning provenance data integration: a practical approach. WWW '23 Companion, page 1542–1550. ACM, 2023.

[29] Débora Pina, Adriane Chapman, Liliane Kunstmann, Daniel de Oliveira, and Marta Mattoso. Dlprov: A data-centric support for deep learning workflow analyses. DEEM '24, page 77–85. ACM, 2024.

[30] Débora Pina, Liliane Kunstmann, et al. Dlprov: a suite of provenance services for deep learning workflow analyses. *PeerJ Comp. Sci.*, 11:e2985, 2025.

[31] Débora Pina, Liliane Kunstmann, Felipe Bevilaqua, Isabela Siqueira, Alan Lyra, Daniel de Oliveira, and Marta Mattoso. Capturing provenance from deep learning applications using keras-prov and colab: a practical approach. *Journal of Information and Data Management*, 13(5), Dec. 2022.

[32] Débora Pina, Liliane Kunstmann, Daniel de Oliveira, and Marta Mattoso. Breadcrumbs for your deep learning model: Following provenance traces with dlprov. *Software Impacts*, 23:100730, 2025.

[33] Débora Pina, Liliane Neves, Daniel de Oliveira, and Marta Mattoso. Captura automática de dados de proveniência de experimentos de aprendizado de máquina com keras-prov. In *Anais Estendidos do XXXVI Simpósio Brasileiro de Bancos de Dados*, pages 69–74, Porto Alegre, RS, Brasil, 2021. SBC.

[34] Jim Pruyne, Justin M. Wozniak, and Ian Foster. Tracking dubious data: Protecting scientific workflows from invalidated experiments. In *2022 IEEE 18th International Conference on e-Science (e-Science)*, pages 456–461, 2022.

[35] Sebastian Schelter, Joos-Hendrik Böse, Johannes Kirschnick, Thoralf Klein, and Stephan Seufert. Automatically tracking metadata and provenance of machine learning experiments. In *Machine Learning Systems workshop at the conference on Neural Information Processing Systems (NIPS)*, 2017.

[36] Marius Schlegel and Kai-Uwe Sattler. Management of machine learning lifecycle artifacts: A survey. *SIGMOD Rec.*, 51(4):18–35, jan 2023.

[37] Marius Schlegel and Kai-Uwe Sattler. Mlflow2prov: Extracting provenance from machine learning experiments. In *Proceedings of the Seventh Workshop on Data Management for End-to-End Machine Learning*, DEEM '23, New York, NY, USA, 2023. Association for Computing Machinery.

[38] Shreya Shankar and Aditya G. Parameswaran. Towards observability for production machine learning pipelines. *Proc. VLDB Endow.*, 15(13):4015–4022, sep 2022.

[39] Filipe Silva, Débora Pina, Liliane Kunstmann, and Marta Mattoso. Painel de proveniência: análise durante o treinamento de redes neurais profundas. In *Anais Estendidos do XXXVI Simpósio Brasileiro de Bancos de Dados*, pages 22–28, Porto Alegre, RS, Brasil, 2021. SBC.

[40] Rômulo Silva, Débora Pina, Liliane Kunstmann, Daniel de Oliveira, Patrick Valduriez, Alvaro Coutinho, and Marta Mattoso. Capturing provenance to improve the model training of pinns: first handon experiences with grid5000. In *42nd Ibero-Latin-American Congress on Computational Methods in Engineering and 3rd Pan American Congress on Computational Mechanics*, pages 1–7, 2021.

[41] Vítor Silva, Daniel de Oliveira, Patrick Valduriez, and Marta Mattoso. Dfanalyzer: runtime dataflow analysis of scientific applications using provenance. *Proc. VLDB Endow.*, 11(12):2082–2085, 2018.

[42] Renan Souza, Leonardo G. Azevedo, Vítor Lourenço, Elton Soares, Raphael Thiago, Rafael Brandão, Daniel Civitarese, Emilio Vital Brazil, Marcio Moreno, Patrick Valduriez, Marta Mattoso, Renato Cerqueira, and Marco A. S. Netto. Workflow provenance in the lifecycle of scientific machine learning. *Concurrency and Computation: Practice and Experience*, 34(14):e6544, 2022.

[43] Renan Souza, Silvina Caino-Lores, Mark Coletti, Tyler J. Skluzacek, Alexandru Costan, Frédéric Suter, Marta Mattoso, and Rafael Ferreira Da Silva. Workflow provenance in the computing continuum for responsible, trustworthy, and energy-efficient ai. In *2024 IEEE 20th International Conference on e-Science (e-Science)*, pages 1–7, 2024.

[44] Jason Tsay, Todd Mummert, Norman Bobroff, Alan Braz, Peter Westerink, and Martin Hirzel. Runway: machine learning model experiment management tool. In *Conference on systems and machine learning (sysML)*, 2018.

[45] Manasi Vartak and Samuel Madden. Modeldb: Opportunities and challenges in managing machine learning models. *IEEE Data Eng. Bull.*, 41(4):16–25, 2018.

[46] Manasi Vartak, Harihar Subramanyam, Wei-En Lee, Srinidhi Viswanathan, Saadiyah Husnoo, Samuel Madden, and Matei Zaharia. Modeldb: a system for machine learning model management. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, pages 1–3, 2016.

[47] Justin M. Wozniak, Zhengchun Liu, Rafael Vescovi, Ryan Chard, Bogdan Nicolae, and Ian Foster. Braid-db: Toward ai-driven science with machine learning provenance. In Jeffrey Nichols, Arthur 'Barney' Maccabe, James Nutaro, Swaroop Pophale, Pravallika Devineni, Theresa Ahearn, and Becky Verastegui, editors, *Driving Scientific and Engineering Discoveries Through the Integration of Experiment, Big Data, and Modeling and Simulation*, pages 247–261, Cham, 2022. Springer International Publishing.

[48] Matei Zaharia, Andrew Chen, Aaron Davidson, Ali Ghodsi, Sue Ann Hong, Andy Konwinski, Siddharth Murching, Tomas Nykodym, Paul Ogilvie, Mani Parkhe, et al. Accelerating the machine learning lifecycle with mlflow. *IEEE Data Eng. Bull.*, 41(4):39–45, 2018.